Quantifying Rhetoric Alignment using Node Embeddings on Temporal Graphs

Keeley Erhardt and Alex Pentland

MIT Media Lab, Cambridge, Massachusetts, USA
{keeley,pentland}@mit.edu

Introduction

In recent years, there has been a growing interest in leveraging node embeddings within graph structures to gain insights into various domains. While text embeddings have been extensively explored for natural language processing tasks, researchers have only recently begun to apply node embeddings in graphs to analyze their underlying structure. This paper utilizes vector representations of a network of Russian and Chinese diplomats' online activity to map the degree of similarity between the diplomats' discourse. This approach extends the application of temporal graph learning, which extracts knowledge from evolving networks, to the domain of political communication and international relations. The analysis provides a deeper understanding of diplomatic interactions within the evolving landscape of digital communication.

Node Embeddings

Node embeddings are low-dimensional representations or vectors that capture the semantic and structural characteristics of nodes within a graph. These embeddings encode information about the relationships, connectivity, and context of nodes. *Node2vec*[1], the most widely used method for node embedding, aims to learn low-dimensional vector representations, denoted as $x_v \in \mathbb{R}^d$, for each node $v \in V$. The algorithm balances the exploration-exploitation trade-off by defining parameters p and q. The parameter p, also known as the return parameter, controls the likelihood of returning to the previous node in a random walk. A higher value of p increases the probability of returning, leading to more localized exploration of the graph. On the other hand, the parameter q, known as the inout parameter, determines the likelihood of visiting nodes that are farther away from the current node in the random walk. A higher value of q favors visiting such nodes, resulting in more global exploration of the graph.

Node2vec maximizes the average log-probability of observing a context node u given a target node v, using the Skip-gram model. The objective function is formulated as maximizing $\sum \log P(u|v)$, where $P(u|v)$ models the probability of observing context node u given target node v. By optimizing this objective function, *node2vec* effectively preserves important graph properties, such as node proximity and structural similarity, in the vector space. In the context of this paper, *node2vec* serves as a powerful tool for extracting meaningful representations of Twitter accounts belonging to Russian and Chinese diplomats, facilitating the analysis of convergence and divergence in their online rhetoric.

Methods

This study investigates the Twitter content shared by 308 Russian diplomats and 222 Chinese diplomats[1] from January 1, 2022, to March 31, 2022. Collectively, these accounts generated 182,693 tweets, which we collected using the Twitter v2 Search API. To examine the temporal evolution of the alignment in online discourse between diplomats from both countries, we construct a graph $G(V,E)$ for each month. This graph captures the dynamic connectivity patterns between the diplomats' online interactions, allowing us to analyze changes in alignment over time. In this graph, V represents the set of Twitter accounts, while E represents the edges connecting pairs of nodes. The edges are determined based on a shared topic criterion, meaning that an edge exists between two nodes if they discuss the same topic. For any two nodes u and v, an edge $(u,v) \in E$ exists if and only if there is a shared topic T that both u and v discuss. To determine these topics, we utilize Twitter's analysis of the tweet's content, returned by the Twitter v2 Search API as context annotations[2].

For each graph G, we learn an embedding over all nodes using `PecanPy`[2], an optimized Python implementation of *node2vec*. We use the `FirstOrderUnweighted` mode with p and q both set to 1. By considering the first-order transition probabilities, the algorithm captures the local neighborhood structure and guides the random walk towards exploring similar or related nodes in the graph. All parameters are set to their default – the dimension of the final embedding is 128, ten random walks are generated from each node, and the length of each random walk is 80.

[1] https://github.com/schliebs/disinfo
[2] https://github.com/twitterdev/twitter-context-annotations

Results

Between January and March 2022, online discussions of Russian and Chinese diplomats reveal a converging pattern. This convergence is supported by a decrease in the Euclidean distance between their respective vector representations. A Euclidean distance of zero would signify complete convergence in the topics addressed by the diplomats. Figure 1 displays the average Euclidean distances, revealing a reduction of over threefold in the average distance between the node embeddings of Russian and Chinese accounts. This intriguing trend suggests evolving alignment in diplomatic rhetoric between the two countries.

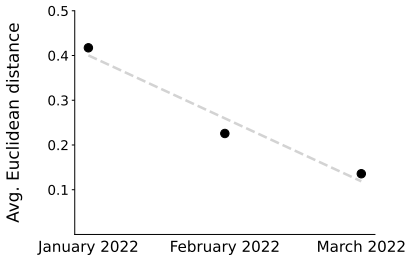

Figure 1. Average Euclidean distance between node embeddings associated with different countries

To provide a visual representation of the node embeddings, we employ t-SNE, a dimensionality reduction algorithm. Figure 2 showcases the visualization of these embeddings in a two-dimensional mapping space. By applying t-SNE, the algorithm learns a mapping from the original high-dimensional vectors to a lower-dimensional space. The resulting mapping ensures that if two high-dimensional vectors, denoted as u and v, are close in proximity, their corresponding mapped points, $map(u)$ and $map(v)$, are also closely positioned within the 2-d mapping space.

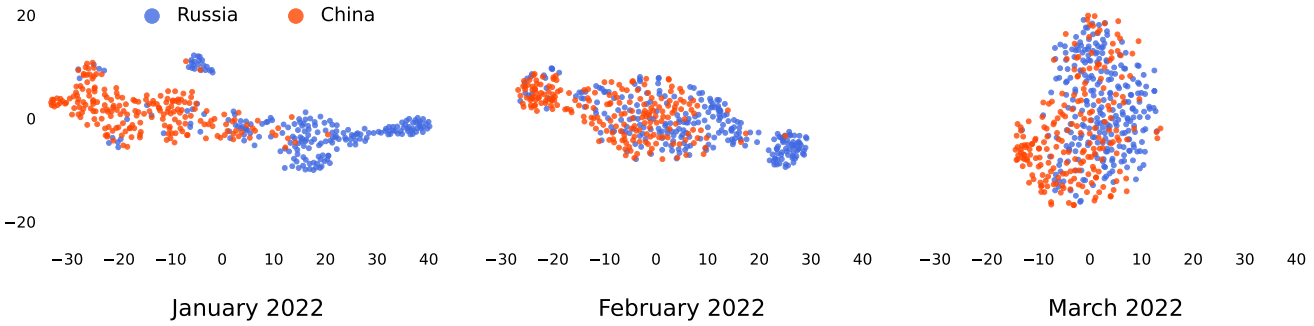

Figure 2. Visualization of node embeddings using t-SNE

In future work, we plan to delve deeper into analysis of the alignment observed between Russian and Chinese diplomats' online discourse. Our focus will be on identifying and examining the specific topics that demonstrate the most notable convergence. This research contributes to a more nuanced understanding of the evolving diplomatic dynamics and signaling strategies employed by both countries, providing valuable insights into their communication patterns and diplomacy efforts.

References

1. Grover, A. & Leskovec, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864 (2016).
2. Liu, R. & Krishnan, A. Pecanpy: a fast, efficient and parallelized python implementation of node2vec. *Bioinformatics* **37**, 3377–3379 (2021).

Acknowledgements

Research was sponsored by the United States Air Force Research Laboratory and the Department of the Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Department of the Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.